

Grid 서비스

김 진 석 | jskim@venus.uos.ac.kr

서울시립대학교 컴퓨터통계학과 교수

오 영 은 | iceize@venus.uos.ac.kr

서울시립대학교 컴퓨터통계학과 석사과정

I. 서 론

II. GridFTP

III. Grid Portal

IV. Grid Information Service

V. Grid Service Working Group

VI. 맷음말

현재 Grid 공동체에서는 여러 종류의 저장 시스템을 사용하고 있다. 그러나 저장 시스템간에 호환성이 유지되지 않아 서로 다른 클라이언트 라이브러리를 사용해야 하는 문제점이 있다. 그래서 저장 장치에 접근하는 방식을 표준화하고자 하여 GridFTP(1)라고 불리는 프로토콜이 제안되었다.

I. 서 론

Grid는 문산회어 있는 고성능 컴퓨팅 자원들을 각자 형태로 둑어 어느 곳에서도 그 자원을 이용할 수 있도록 해주는 서비스이다. 그러나 이러한 Grid 환경을 구축하기 위해서는 해결해야 할 과제가 많다.

본고에서는 Grid 서비스 관련 기술에 대하여 고찰한다. 즉, 대량의 데이터 전송을 위한 GridFTP, 사용자가 웹을 통하여 쉽게 자원들을 이용할 수 있게 해주는 Grid Portal, Grid 환경에서의 자원의 위치, 상세 정보 등을 구축하는 Grid 정보 서비스에 대하여 알아본다.

II. Grid FTP

대량의 데이터를 사용하는 과학 및 공학 용용 등은

분석과 시각화를 위해 내용당의 저장 시스템과 고속의 네트워크 필요하다. 그리고 이러한 용용에서는 지리적으로 분산된 많은 용용 프로그램에 의한 데이터 접근이 요구된다. 현재 Grid 공동체에서는 여러 종류의 저장 시스템을 사용하고 있다.

그러나 저장 시스템간에 호환성이 유지되지 않아 서로 다른 클라이언트 라이브러리를 사용해야 하는 문제점이 있다. 그래서 저장 장치에 접근하는 방식을 표준화하고자 하여 GridFTP(1)라고 불리는 프로토콜이 제안되었다. GridFTP는 Grid 데이터 전송 및 접근 프로토콜로서 Grid 환경에서 안정적이고 효과적인 데이터 전송을 제공한다.

GridFTP는 FTP 프로토콜을 확장한 것이다. 다음 절에서는 GridFTP의 특징, 구현 및 성능에 대해 간략하게 살펴보자 한다.

미들웨어

1. GridFTP의 특징

GridFTP는 다음과 같은 특징이 있다.

- 1) GSI(Grid Security Infrastructure)와 Kerberos를 지원한다

GridFTP는 RFC 2228 "FTP Security Extensions"에 정의된 GSSAPI (Generic Security Service Application Programming Interface) 인증 방식을 구현함으로써 파일을 접근하거나 전송할 때 확고하고 유통성 있는 인증, 무결성, 그리고 기밀성을 제공한다.

2) Third-party 챈들에 의한 데이터 전송을 지원한다
분산되어 있는 많은 데이터를 관리하기 위해서는 저장 시스템간의 데이터 전송을 위하여 인증된 Third-party에 의한 챈들 제공해야 한다. 이러한 동작들은 중신지와 수신지 간의 데이터 전송의 초기화, 모니터링, 챈들 제공하며, 인증을 위하여 GSSAPI를 사용한다

- 3) 병렬 데이터 전송을 지원한다

광역 네트워크 망에서는 데이터를 병렬로 전송하는 것이 하나로 전송하는 것보다 대역폭을 더 향상할 수 있다. GridFTP는 FTP 명령어의 확장과 데이터 채널 확장을 통하여 이러한 병렬 데이터 전송을 제공한다.

- 4) 헤어지거나 중복된 데이터 전송을 지원한다

이것은 여러 서버에 문화되어 있는 데이터를 여러 개의 TCP 스트림을 사용하여 전송하며 병렬 전송으로 대역폭을 향상시킨다.

- 5) 부분 파일 전송을 지원한다

많은 응용 프로그램에서는 전체 파일보다 파일의 부분만을 전송하는 것이 더 좋다. 이것은 큰 객체지향 물리 데이터베이스에서 일부분만을 접근하는 고에너지 물리 분석 등과 같은 응용에서는 중요하다.

- 6) TCP 버퍼와 창(window)의 크기를 자동으로 조절한다

TCP 버퍼와 창의 크기를 최적으로 조정하는 것은 데이터 전송 성능에 지대한 향상을 기할 수 있지만,

수동으로 조정하게 되면 에러가 발생할 수 있다. GridFTP는 용량이 큰 파일 및 작은 파일의 묶음을 등을 전송하기 위해 자동과 수동으로 TCP 버퍼 크기를 조정한다.

- 7) 신뢰할 수 있는 데이터 전송을 제공한다

네트워크 상의 문제, 서버의 정전으로 인하여 전송이 중단되었을 경우, 이를 재전송할 수 있도록 해준다.

2. GridFTP의 구현

Globus 컴퓨팅 환경에서 구현된 GridFTP는 현재까지 부분적인 파일 전송, 제3자에 의한 전송, 병렬 전송, 헤어지거나 중복된 데이터 전송을 지원하며, 아직 TCP 버퍼와 창의 크기의 자동적인 조정 기능을 제공하지 않는다.

GridFTP는 C언어로 구현된 두 개의 주 라이브러리인 `globus_ftp_control_library`와 `globus_ftp_client_library`로 구성되어 있다^[5]. `globus_ftp_control_library`는 API 챈들 채널을 제공한다. 이 API를 이용하여 인증, 연결 관리, 챈들 채널 및 데이터 채널의 생성, 그리고 데이터 채널에서의 읽기, 쓰기를 제공한다. 또한 병렬 전송, 헤어지거나 중복된 데이터 전송, 그리고 제3자에 의한 전송과 같은 특징을 지원한다. `globus_ftp_client_library`는 GridFTP 클라이언트 API를 제공한다. 현재 Globus 블켓 2.0 베타 버전^[6]에서는 GridFTP를 위하여 `gsiftp` 페키지를 제공하고 있으며, FTP 서버는 `wu-ftpd`, 클라이언트는 `ncftp`를 수용하여 사용하고 있다.

3. GridFTP의 성능

〈그림 1〉은 동시에 전송하는 TCP 스트림의 수를 1부터 32까지 증가했을 때 GridFTP의 병렬 전송의 성능을 보여준다. 그러나 스트림의 수가 증가함에 따라 그에 따른 이익은 줄어든다. 이 테스트에서 8개의 스트림을 사용하는 것이 병렬 전송에 가장 좋다는 것을 알 수 있다.

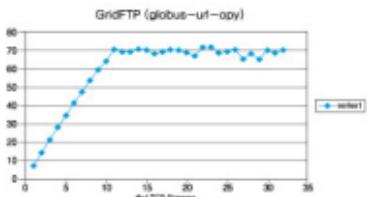


그림 1) TCP 스트림의 수를 증가시킬 때의 병렬 전송 대역폭¹⁾

GridFTP의 다음 버전에는 데이터 채널을 캐싱하여 더 높은 대역폭을 지원할 것이다. 이렇게 되면 전송이 종료되자마자 전송에 사용된 TCP 스트림을 바로 사용하여 각각의 전송 사이에 TCP의 연결 해제와 재전송에 걸리는 시간을 없앨 수 있다.

III. Grid Portal

Grid portal은 사용자 작업을 수행하기 위하여 Grid 서비스 및 자원을 활용하는 도구를 제공하는 것뿐만 아니라 Grid 서비스에 대한 정보를 안전하게 수집할 수 있도록 해 주는 웹 용용 서버의 한 부류이다²⁾. 이렇게 하려면 반드시 사용 가능한 커포넌트 개발이 필수적이며, 다양한 요구에 맞게 커포넌트를 이용하여 portal을 구성하게 된다. 이러한 커포넌트는 반드시 쉬운 협업, 작업 의뢰, 작업 감시, 커포넌트 발견, 영구적인 객체 저장 등의 기능을 제공해야 한다³⁾. 그리고 이러한 기능을 구현한 GPKD(Grid Portal Development Kit)은 다음과 같은 구조로 되어 있다⁴⁾.

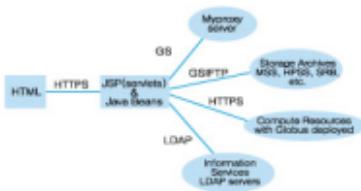


그림 2) Grid Portal Development Kit의 구조

그림 2에서 사용자는 웹 브라우저를 통해 웹 서버에 안전하게 접근하게 된다. 웹 서버에서는 JSP와 Java Beans를 이용하여 Myproxy를 통한 사용자 인증과 GSI/FTP를 통한 파일 전송, LDAP(Lightweight Directory Access Protocol)를 통한 정보 발견을 하게 되며 이와 같은 과정을 통하여 계산을 수행하게 된다⁵⁾. 본 고에서는 Grid Portal을 개발하거나 운영하고 있는 실제 사이트의 예를 살펴본다.

1. Grid Portal Collaboration

NASA에서 주관하는 IPG(Information Power Grid)에는 Grid에 접근하는 데 필요한 기술을 개발하기 위한 Grid portal 공동 연구가 진행되고 있다⁶⁾. NCSA(National Center for Supercomputing Applications), SDSC(San Diego Supercomputer Center), NASA IPC에서는 일반적인 형태의 portal 구조를 개발하고 있으며, 모든 사이트에서 사용할 수 있는 재사용 가능한 커포넌트를 제공하고 있다. 현재는 원격 자원에 안전하게 접근할 수 있도록 Globus 블켓을 미들웨어로 사용하고 있다.

2. Cactus Portal Grid

Cactus⁷⁾은 과학, 공학 문제를 풀기 위한 공개된 소스를 제공하고 있으며, Globus 블켓, HDF5 파일 입출력, PETSc 과학 라이브러리, 고급 영상 도구 등을 포함하는 소프트웨어 기술들을 쉽게 접근할 수 있게 한다.

3. NPACI Grid Portal

NPACI(National Science Foundation's Partnerships for Advanced Computational Infrastructure)에서 제공하는 Grid portal은 사용자로 하여금 컴퓨팅 자원을 웹 페이지상으로 손쉽게 접근할 수 있도록 한다. 이 portal은 GridPort 블켓⁸⁾을 기반으로 제작되

미들웨어

있으며, 사용자가 수퍼컴퓨터의 계정과 작업 관리 도구, 파일 시스템, 시스템의 전반적인 정보에 대하여 직접 접근할 수 있게 한다.

(그림 3)은 PACI의 Grid portal 시스템의 상태를 보여주고 있다. 이 그림을 보면 특정 사이트에 어떤 종류의 컴퓨터가 동작하고 있으며 그 컴퓨터의 부하 및 규모의 상태를 알 수 있다.

IMPACT Computer and Network Systems								
System	Category	Type	Computer Systems					
			Name	Status	Level	Planning	Demand	Current
UCSC	UCSC	UCSC-1000	host1	active	0	70%	9	14
UCSC	UCSC	UCSC-1000	host2	active	0	85%	9	15
UCSC	UCSC	UCSC-1000	host3	active	0	65%	20	4
UCSC	UCSC	UCSC-1000	host4	active	0	50%	4	10
UCSC	UCSC	UCSC-1000	host5	active	0	12%	3	8
UCSC	UCSC	UCSC-1000	host6	idle	0	75%	0	0
UCSC	UCSC	UCSC-1000	host7	active	0	60%	11	22
UCSC	UCSC	UCSC-1000	host8	idle	0	80%	0	0
UCSC	UCSC	UCSC-1000	host9	idle	0	55%	14	8
UCSC	UCSC	UCSC-1000	host10	idle	0	45%	29	35
UCSC	UCSC	UCSC-1000	host11	idle	0	12%	30	33
UCSC	UCSC	UCSC-1000	host12	idle	0	25%	32	36
UCSC	UCSC	UCSC-1000	host13	idle	0	18%	3	8
UCSC	UCSC	UCSC-1000	host14	idle	0	10%	0	0
UCSC	UCSC	UCSC-1000	host15	idle	0	70%	0	0
UCSC	UCSC	UCSC-1000	host16	idle	0	50%	2	2
UCSC	UCSC	UCSC-1000	host17	idle	0	30%	9	14

〈그림 3〉 Grid portal AI스냅

4. Legion Grid Portal

Legion에서 제공하는 Grid portal¹¹⁾은 portal을 이용하여 사용자가 파일의 내용, 작업의 수행과 감시, 계정 정보 등을 직관적으로 볼 수 있게 해준다. 현재는 Perl Script, C/C++, PHP 언어와 MySQL 데이터베이스를 사용하여 구현되어 있다.

IV. Grid Information Service

GIS(Grid Information Service)는 Grid에 관련된 정보를 제공하거나 사용하는 분야이다. 분산 컴퓨팅 환경에서의 고성능 연산을 할 때에는 컴퓨터, 네트워크 및 기타 자원 분 아니라 유통에서 사용하는 프로토콜이나 알고리즘을 주의하여 선택하는 것이 필요하다. 이러한 선택을 하는데 가장 필수적인 요소는 사용 가능한 자원의 상태와 구조에 대한 정확하고 최근의 정보를 제공하는 일이다^[4]. I-WAY는 대량의 병렬 컴퓨터, 워크스테이션, 저장 시스템 그리고 시각, 청각 등으로 구성되어 있는

다¹⁰. 이러한 자원들은 인터넷과 155Mbps 전용 ATM 네트워크에 서로 연결되어 있다. 이 환경에서 웹 프로그램들은 다양한 환경에서 동작할 수 있다. 그러므로 최적의 설정을 찾기 위해서는 자원들에 대한 상세한 정보를 가지고 있어야 한다. 예를 들어, ATM 네트워크는 현재 사용 가능한가, 혹은 반드시 TCP/IP를 사용해야 하는 노드들은 어떤 것들이 있는가에 대한 질문에 대해서 답할 수 있어야 한다. 이러한 문제를 해결하기 위해서는 MDS(Metacomputing Directory Service)를 경의하고 구현되었다. MDS는 LDAP을 기반으로 하여 여러 자원에 대한 정보를 하나의 형태로 쉽게 찾을 수 있도록 하는 API(Application Programming Interface)이다. 또한 Legion¹¹, NEOS¹², NetSolve¹³, Condor¹⁴, Nimrod¹⁵, PRM¹⁶, AppleS¹⁷ 환경에서 유용하게 사용될 수 있으며, MPI¹⁸로 연결된 이기종 분산 시스템에서도 사용될 수 있다.

1. MDS의 개요

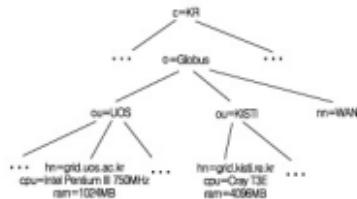
LDAP은 모든 정보를 object class를 이용하여 표시하고 있다. MDS도 이와 같이 자원에 대한 상세한 정보를 object class로 표현하고 있는데, 이는 X.500의 방식을 따른 것이다. 각각의 MDS 엔트리들은 DN(Distinguished Name)이라고 하는 것으로 구분된다. 예를 들면 호스트이름이 grid.ucs.ac.kr이고 기관명이 서울시립대학교, 사용자명이 Oh일 때 DN은 다음과 같다. 이 때 hn은 hostname, ou는 organization unit, cn은 common name을 나타낸다.

hn = grid nos ac kr

OU = University of Seoul

$\alpha = 0$

또한 각각의 MDS 엔트리들이 어디에 위치하고 있는지 표현하기 위해 object class는 개별 구조로 이루어져 있다.



〈그림 4〉 노드와 네트워크 정보를 MOS 계층 구조로 표현한 예

<그림 4>에서는 서울시립대학교(UOS)에 있는 노드와 한국과학기술정보연구원(KISTI)에 있는 노드가 WAN으로 연결되어 있다는 것과 각각의 노드가 갖는 CPU와 메모리 상태를 표현하고 있다.

2. MDS의 구원

MDS에서는 네트워크 표현하기 위해서 RFC 1609²²을 기반으로 구성하였고 몇 가지 문제점을 해결하기 위해서 object class를 확장하였다²³. (그림 5)는 새로 추가된 MDS object class의 간단한 형태를 보여주고 있다.

```

G:\junction\OBJET Class          G:\junction\OBJET Class
  SURFACE OF Substances        SURFACE OF top
  MUST CONTAIN {               MUST CONTAIN {
    bottom:                   }           }
    type:                     manager:      11 obj
    width:                    previous:    11 obj
    note:                     technician: 11 obj
    subtype:                  description: 11 obj
    dimension:                documentation: 11 obj
  }
  MAY CONTAIN {               }
    carbohydrate:            11 obj
    totalEnergy:              11 obj
    totalSwap:                11 obj
    dataSeries:               11 obj
    instructionCode:         11 obj

```

〈그림 5〉 MDS object class의 구조화한 형태

MDS에서는 〈그림 5〉과 같은 형태로 정보들을 저장하게 되며, 이 정보들을 가져오는 단일화된 방법을 제공하고 있다. 현재 Globus 블켓에 구현된 MDS의 object class 구조²²⁾를 살펴보면, CPU, 메모리, 파일 시스템, 네트워크 MDS에 대한 정보를 저장하고 있음을

알 수 있다.

〈그림 6〉은 Globus 블켓에서 제공하는 grid-search 명령어를 이용하여 grid.ucs.ac.kr이라는 호스트에서 정보를 가져온 결과를 보여준다. 이 그림에서 첫째줄은 DN을 표시하고 있는 것을 알 수 있다. 또한 호스트 컴퓨터는 PentiumIII 933MHz의 듀얼 CPU를 사용하며 각각의 CPU마다 256KByte의 내부 캐시를 가지고 있다는 정보를 보여주고 있다.

〈그림 6〉 grid-item-position을 설정한 결과

V. Grid Service Working Group

GGF(Global Grid Forum) Grid Service 위킹그룹은 Grid를 기반으로 한 컴퓨팅을 하는 데 필요한 정보 서비스의 개발 및 요구 사항을 명확히 하는 일을 수행한다. 이 위킹그룹에서는 GOS(Grid Object Specification), GNF(Grid Notification Framework), MDSI(Metacomputing Directory Service)의 세 분야를 다룬다. GOS는 정보 서비스의 DDL(Data Definition Language)이며 Grid 기반 응용 프로그램이나 Grid 환경의 자원들을 간단하게 기술할 수 있게 한다. 그리고 이렇게 만들어진 GOS 포맷은 MDS에서 사용하여 아주 간단하고 일반적인 형태로 구현하여

미들웨어

ASN.1, LDAP, XML, HTML, SQL 데이터와 같은 여러 형태로 변역할 수 있게 된다^[2]. GNF는 GOS로 설명되는 Grid 자원이 실제로 있는지 발견하는 역할을 한다. 이 프레임워크는 Grid 자원이 주기적으로 메시지를 주고받으며 항상 최근의 상태를 유지하기 위해 TS(Time Stamp)를 사용한다^[2]. 마지막으로 MDS는 Grid 환경에서의 자원의 위치 및 상세 정보 등을 알려주는 서비스이다.

VI. 링크말

최근 여러 컴퓨터 지원을 병렬로 결합하여 대규모 컴퓨터 시스템을 구현하는 Grid 기술에 대한 연구가 활발하게 진행되고 있다. Grid 시스템을 이용하면 기존에는 해결하기 어려웠던 대규모 병렬/분산 처리 문제들을 해결할 수 있다. 본고에서는 대량의 데이터를 병렬로 전송할 수 있는 GridFTP 사용자가 맵으로 접근하여 수퍼컴퓨터의 상태를 확인하고 이를 쉽게 이용할 수 있게 해주는 Grid portal, Grid 환경에서 자원의 정보를 제공하여 주는 MDS 기술에 대하여 살펴보았다.

■ 참고문헌

- [1] B. Alcock, J. Becker, J. Braemmer, A. Chervenak, I. Foster, C. Kesselman, S. Mader, V. Nedelova, D. Quigley, and S. Tuecke, "Secure, Efficient Data Transport and Replica Management for High-Performance Data-Intensive Computing," Proc. of IEEE Mass Storage Conference, 2001.
- [2] GridFTP: <http://www.globus.org/distributed/grid/GridFTP.pdf>
- [3] Globus GridFTP Tools: <http://www.globus.org/gftp/applications/tools.html>
- [4] Globus: <http://www.globus.org>
- [5] Grid Portal: http://www.gridco.de/Präsentation/GridPortal_2000.ppt
- [6] G. von Leszczewski and I. Foster, "Grid Infrastructure to Support Science Portals for Large Scale Instruments," Proc. of the Workshop Distributed Computing on the Web, 1999.
- [7] Grid Portal Development Kit: <http://www-tp.bipr.gov/Grid/projects/GPK/>
- [8] Grid Computing Environments Community Practice: <http://www.ggf.ni/activities/GCE/GCE-GPK-BP.pdf>
- [9] Portal Collaboration: <http://www.egnss.gov/portals/intro.html>
- [10] Gnosis Code: <http://www.gnosiscode.org/Documentation/introduction.html>
- [11] M. Thomas, S. Mock, and J. Boisseau, "Development of Web Tools for Computational Science Portals: The NPACI HotPage," Proc. IEEE Symp. on High-Performance Distributed Computing, 2000.
- [12] PACI Homepage: <http://paci.eeic.org>.
- [13] A. Natalej, A. Nguem-Tsing, M. Humphrey, and A. Grimshaw, "The Legion Grid Portal," Grid Computing Environments, 2001.
- [14] S. Fitzgerald, I.-Kozai, C. Kesselman, G. von Leszczewski, W. Smits, and S. Tuecke, "A Directory Service for Configuring High-Performance Distributed Computation," Proc. 6th IEEE Symp. on High-Performance Distributed Computing, 1997.
- [15] T. Dufresne, I. Foster, M. Papka, R. Stevens, and T. Kuhlwein, "Overview of the i-WAY: Wide Area Visual Supercomputing," International Journal of Supercomputer Applications, 1996.
- [16] A. Grimshaw, J. Weesener, and E. Lyot, "Jr. Metasystems: An Approach Combining Parallel Processing and Heterogeneous Distributed Computing Systems," Proc. 6th International Conference on Parallel and Distributed Computing, 1994.
- [17] J. Czylik, M. Meier, and J. Mir, "The Network-Enabled Optimization System (NEOS) Server," Preprint MCS-P615-0996, Argonne National Laboratory, 1996.
- [18] H. Casanova and J. Dongarra, "NetSolve: A Network Server for Solving Computational Science Problems," Technical Report CS-95-93, University of Tennessee, 1995.
- [19] M. Liskow, M. Livny, and M. Muko, "Condor - A Hub of the Workstations," Proc. 8th International Conference on Distributed Computing Systems, pp. 104-111, 1988.
- [20] D. Abramson, R. Sosik, J. Giddy, and B. Hill, "Nimrod: A Tool for Performing Parameterized Simulations Using Distributed Workstations," Proc. 4th IEEE Symp. on High Performance Distributed Computing, 1995.
- [21] B. Neumann and S. Rao, "The Propriero Resource Manager: A Scalable Framework for Processor Allocation in Distributed Systems," Concurrency: Practice & Experience, vol. 8, no. 4, pp. 39-65, 1994.
- [22] F. Berman, R. Wolski, S. Figueira, J. Schopf, and G. Sheo, "Application-Level Scheduling on Distributed Heterogeneous Networks," Proc. of ACM Supercomputing, 1996.
- [23] W. Gropp, E. Lusk, N. Doss, and A. Skjellum, "A High-Performance, Portable Implementation of the MPI Message Passing Interface Standard," Parallel Computing, vol. 22, pp. 769-808, 1996.
- [24] I. Foster and G. von Leszczewski, "Usage of LDAP in Globus".
- [25] G. Marshall, T. Johnson, and M. Krueper, "Sharing networks in the x500 directory," RFC-1639, 1994.
- [26] IETF 2.1 schema: <http://www.ietf.org/rfc/rfc2256.txt>
- [27] IGS Working Group: <http://www-uniserv.unisys.jp/igw/igw/igs/>
- [28] GDS: <http://www.unisys.jp/gdws/gdws/gs/gs-wg-021-000.html>
- [29] gids: <http://www-tp.bipr.gov/GridForum/gids.html>